Virologica Sinica (2020) 35:388–397
https://doi.org/10.1007/s12250-020-00194-6

www.virosin.org
www.springer.com/12250

RESEARCH ARTICLE

Conservative Evolution of Hepatitis B Virus Precore and Core Gene During Immune Tolerant Phase in Intrafamilial Transmission

Yuqian Luo[1] · Le Zhang[1] · Yimin Dai[2] · Yali Hu[2] · Biyun Xu[3] · Yi-Hua Zhou[1,4]

Received: 2 August 2019 / Accepted: 6 December 2019 / Published online: 2 March 2020
© Wuhan Institute of Virology, CAS 2020

Abstract

Hepatitis B virus (HBV) is characterized with high mutations, which is attributed to the lack of proof-reading of the viral reverse transcriptase and host immune pressure. In this study, 31 HBV chronic carriers from 14 families were enrolled to investigate the evolution of the same original HBV sources in different hosts. Sequences of pre-C and C (*pre-C/C*) genes were analyzed in eight pairs of HBV-infected mothers with longitudinal sera (at an interval of 6.0–7.2 years) and their children (5.5–6.7 years old), and in 15 adults (21–78 years old) from six families with known intrafamilial HBV infection. The *pre-C/C* sequences had almost no change in eight mothers during 6.0–7.2 years and their children who were in immune tolerant phase. The *pre-C/C* sequences from the 15 adults of six families, mostly in the immune-clearance phase or the low replicative phase, showed various diversified mutations between individuals from each family. Compared to a reference stain (GQ205441) isolated nearby, the *pre-C/C* in individuals in immune tolerant phase showed 98.56%–99.52% homology at nucleotide level and 99.5%–100% homology at amino acid level. In contrast, multiple mutations were developed in the immune-clearance phase or the low replicative phase, affecting immune epitopes in *core* gene and G1896 in *pre-C* gene. The results indicate that the evolution of new HBV variants is not mainly resulted from the spontaneous error rate of viral reverse transcription, but from the host immune pressure.

Keywords Hepatitis B virus (HBV) · *Pre-C/C* gene · Sequence evolution

Introduction

Chronic hepatitis B virus (HBV) infection is a worldwide health problem. HBV is an enveloped virus that contains a circular DNA genome, approximately 3.2 kb, including four overlapping open reading frames (*S*, *C*, *P*, and *X* genes), and replicates its genomic DNA through reverse transcription by the viral reverse transcriptase (Valaydon and Locarnini 2017). The hepatitis B core antigen (HBcAg) and hepatitis B surface antigen (HBsAg) are major targets for antiviral immunity, but the first one seems to be the most immunogenic (Vanlandschoot *et al.* 2003) and subjected to a wide variation of amino acid sequences during chronic HBV infection (Carman *et al.* 1995). Indeed, it is estimated that the mutation rate is 10^{-4}–10^{-6} nucleotide substitutions/site/year in HBV *pre-C/C* gene, approximately 100 times higher than that of any other DNA virus (Girones and Miller 1989). Lack of proof-reading of the viral reverse transcriptase (Valaydon and Locarnini 2017) and host immune pressure (Wang *et al.* 2010) are reasoned to be the basis for the high mutation rate of HBV.

Yuqian Luo and Le Zhang have contributed equally to this work.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s12250-020-00194-6) contains supplementary material, which is available to authorized users.

✉ Yi-Hua Zhou
zgr03summer@126.com

[1] Department of Laboratory Medicine, Nanjing Drum Tower Hospital and Jiangsu Key Laboratory for Molecular Medicine, Nanjing University Medical School, Nanjing 210008, China

[2] Department of Obstetrics and Gynecology, Nanjing Drum Tower Hospital, Nanjing University Medical School, Nanjing 210008, China

[3] Department of Biostatistics, Nanjing Drum Tower Hospital, Nanjing University Medical School, Nanjing 210008, China

[4] Department of Infectious Diseases, Nanjing Drum Tower Hospital, Nanjing Medical University, Nanjing 210008, China

However, investigative tools such as cell culture systems that support long-term HBV propagation are still lacking.

Dictated by the host immunity, the natural course of chronic HBV infection may undergo five phases: the immune tolerant phase (HBeAg-positive chronic infection), the immune-clearance phase (HBeAg-positive chronic hepatitis), the low replicative phase (HBeAg-negative chronic infection), the immune reactivation phase (HBeAg-negative chronic hepatitis), and, rarely, the recovery (HBsAg-negative phase) (EASL 2017). Patients with mother-to-infant transmission of HBV often remain in the immune tolerant phase for decades. These patients have very weak specific T cell responses, as the HBV DNA levels are high and constant, alanine transaminase (ALT) values are normal, and there is minimal liver inflammation (EASL 2017). Later they may enter into the immune-clearance phase, during which the host immunity is, for unknown reason, awakened to cause notable liver inflammation. The immune-clearance phase may then transit into the low replicative phase, during which seroconversion to positive anti-HBe sometimes occurs, HBV DNA remains at very low levels, and ALT return to normal; alternatively some patients enter into the reactivation phase with higher ALT and return of viremia.

The key to tackle HBV infection seems to lie on the interaction between the host immunity and the virus adaptation. Given the same source of infection, mother-to-infant or intrafamilial transmission of HBV provides a useful model to characterize the virus evolution pattern under different host immune background. In the present study, to shed light on the interaction between host immunity and HBV evolution, we performed longitudinal analysis of HBV *pre-C/C* sequences within family members infected with the same original pathogen source.

Materials and Methods

Study Population

The subjects enrolled in this study contained two groups of individuals with ongoing HBV infection. Group one included eight pairs of relatively young mothers (aged 28–37 years) and their children (aged 5.5–6.7 years) (Hu *et al.* 2012). During a follow-up period of 6.0–7.2 years, these mothers' health conditions were generally well, with normal liver function and normal ultrasound B liver scanning image, and they were positive for both HBsAg and HBeAg and had constantly high HBV DNA levels. Thus, they were in HBeAg-positive chronic infection phase of infection. The longitudinal blood samples from these mothers, collected at an interval of 6.0–7.2 years (Hu *et al.* 2012), were kept at − 20 °C. Their eight children received hepatitis B immunoglobulin and/or three doses of hepatitis B vaccine on a standard 0-, 1-, and 6-month schedule after birth, but they were infected with HBV; the infection was defined in three children before 3 years age (Table 1: Ib, Vb, and XIIIb) and in five others at 5.6–6.7 years age. These children were considered be perinatally infected with HBV as a consequence of immunoprophylaxis failure. The blood samples of these children were only retained at the age of 5.5–6.7 years.

Group two contained 15 individuals from six families, including six index patients (aged 40–78 years), and nine patients (aged 21–53 years) who were assumed to have acquired the infection in their early childhood from the index, except in family XIV who were two spouses, and the husband was assumed to have acquired the infection from his wife as he was negative for HBV before he got married.

Of the above 31 patients, 20 (64.5%) were women. All the patients had no co-infection of HIV or hepatitis C virus, and had not been treated with antiviral agents. The demographic data of these patients are presented in Tables 1 and 2.

Detection of Serological Markers for HBV Infection

Serum samples were tested for HBsAg, antibody against HBsAg (anti-HBs), HBeAg, anti-HBe, and anti-HBc using enzyme-linked immunosorbent assay kits (Huakang Biotech, Shenzhen, China). Quantification of serum HBsAg and HBeAg was performed by a microparticle enzyme immunoassay (Architect System, Abbott, North Chicago, IL, USA), as previously reported (Liu *et al.* 2015). HBeAg levels were presented as the ratio of relative light units of the samples to negative controls (S/CO). Quantification of HBV DNA was performed by a fluorescent real-time PCR assay (Shenyou Biotechnology, Shanghai, China) as described before (Liu *et al.* 2015).

HBV Genotyping and Analysis of *Pre-C/C* Mutations

Serum DNA was extracted from 200 μL serum by phenol/chloroform extraction method, and dissolved in 20 μL Tris–EDTA buffer as reported previously (Xu *et al.* 2017). The *pre-C/C* regions were amplified by nested PCR using primers as listed in Supplementary Table S1. The first round PCR was carried out using primers C1 and C2. The second round was performed using C3 and C4 primers. To prevent cross contamination, each step was performed in separate areas with dedicated equipment, and always included negative controls.

The purified PCR products were directly sequenced on an ABI Prism 3730xl sequencer (Applied Biosystems,

Table 1 Virological characteristics in group one consisting of younger mothers and their children.

Family	Patient	Relation	Sex	Age (years)	HBsAg (IU/mL)	HBeAg (S/CO)	HBV DNA (IU/mL)
I	Ia	Mother	F	30.7	128,640	2299.3	1.42×10^8
				37.0	256,360	2928.0	5.50×10^7
	Ib	Daughter	F	5.9	70,720	2055.2	1.14×10^8
II	IIa	Mother	F	21.4	52,890	2607.8	9.04×10^6
				28.0	29,124	2119.4	3.00×10^6
	IIb	Daughter	F	6.1	57,550	1719.4	8.99×10^6
III	IIIa	Mother	F	25.5	3448	3270.6	5.91×10^6
				32.3	8752	1885.7	5.29×10^6
	IIIb	Daughter	F	6.4	11,130	1876.5	5.97×10^6
IV	IVa	Mother	F	25.8	149,280	2297.9	1.06×10^7
				33.0	139,260	2794.2	1.08×10^7
	IVb	Son	M	6.7	133,200	2178.9	2.32×10^7
V	Va	Mother	F	25.6	36,758	1820.9	2.39×10^6
				32.7	94,220	2612.9	3.67×10^7
	Vb	Son	M	6.7	31,676	1662.9	4.11×10^7
VI	VIa	Mother	F	22.5	32,580	1542.3	5.99×10^6
				28.5	81,490	1829.3	2.63×10^7
	VIb	Daughter	F	5.6	60,600	2279.6	1.27×10^8
VII	VIIa	Mother	F	27.4	37,688	1133.0	5.68×10^6
				34.3	60,530	1510.8	1.27×10^7
	VIIb	Daughter	F	6.5	69,900	1693.3	2.14×10^7
VIII	VIIIa	Mother	F	24.2	84,270	1655.5	2.16×10^7
				30.2	27,038	1759.0	2.60×10^7
	VIIIb	Son	M	5.5	70,942	1489.8	2.13×10^7

All patients were negative for IgM antibody against hepatitis B core antigen, and had normal levels of ALT. They were in the phase 1, the HBeAg-positive chronic infection phase, previously known as the immune tolerant phase of HBV infection.

Hitachi, Tokyo, Japan) after reaction with BigDye Terminator v3.1 (Applied Biosystems, Foster, CA, USA). When mixed signals (multiple peaks) were seen in the chromatograms of sequencing results, the PCR products were subcloned using pUCm-T vector (Sangon Biotech, Shanghai, China).

HBV genotypes were determined by phylogenetic analysis as previously described (Luo *et al.* 2019) based on the *pre-C/C* sequence. Sequences were also aligned with reference stains, including one (GQ205441) isolated in nearby city Hefei in eastern China (Zhang *et al.* 2011), one (KR013798) in Guangzhou in southern China (Liang *et al.* 2015), one (KU519422) in Tibet in western China, and one (LC170476) in Japanese patients. Multiple sequence alignments were performed using Clustal W method. Phylogenetic trees were constructed using neighboring-joining methods (Saitou and Nei 1987) with pairwise distances being estimated by Kimura's two-parameter method (Kimura 1980). The evolutionary distances were computed using the maximum composite likelihood method (Tamura *et al.* 2004). These analyses were done automatically using MegAlign software program (Clewley and Arnold 1997).

Statistical Analysis

Statistical analysis was performed with SPSS software (SPSS Standard v. 17.0, Chicago, IL). Unpaired t test was used to determine the significance in mutation rates between two groups with 95% confidence intervals (CI). All tests were two-sided; $P < 0.05$ was considered as a significant difference.

Results

Status of Hepatitis B Serological Markers and HBV DNA Level

As shown in Table 1, the serum levels of HBsAg, HBeAg, and HBV DNA in each of eight mothers in group one had no significant changes during 6.0–7.2 years (from the second trimester of pregnancy to 5.5–6.7 years postpartum). The HBV DNA levels were constantly higher than 1.0×10^6 IU/mL among all eight mothers. Meanwhile, the levels of HBsAg, HBeAg, and HBV DNA in each child

Table 2 Virological characteristics in group two consisting of elderly parents and adult offspring.

Family no.	Patient no.	Relation	Sex	Age (Years)	HBsAg (IU/mL)	HBeAg (S/CO)	anti-HBe (S/CO)	HBV DNA (IU/mL)	ALT (U/L)	Phase[a]
IX	IXa	Index	F	78	92	0.4	0.2	5.51×10^6	969.8	2
	IXb	Daughter	F	53	1803	–	+	2.40×10^2	12.6	3
	IXc	Son	M	45	97,005	2214.9	–	1.59×10^8	34.5	1
X	Xa	Index	F	71	254	–	+	2.86×10^6	237.1	4
	Xb	Son	M	40	168	–	–	1.32×10^2	16.9	3
XI	XIa	Index	F	63	244.11	1.58	1.72	Undetectable	78.2	2
	XIb	Son	M	41	983	6.9	1.75	1.05×10^6	44.8	2
	XIc	Son	M	38	827	–	+	3.28×10^2	17.9	3
XII	XIIa	Index[b]	F	46	–	–	+	Undetectable	22.4	5
	XIIb	Daughter	F	23	10,182	–	+	3.40×10^2	24.6	3
	XIIc	Son	M	21	4678	490.4	16.02	1.26×10^6	550.3	2
XIII	XIIIa	Index	M	46	5725	–	+	1.51×10^2	40.3	3
	XIIIb	Son	M	22	5202	–	+	1.43×10^2	19.5	3
XIV	XIVa	Index	F	40	155	0.6	+	1.94×10^3	71.8	2
	XIVb	Husband	M	40	0.8	–	–	1.25×10^2	14.7	3

All patients were positive for total antibody against hepatitis B core antigen.

[a]Phases 1–5 refer to HBeAg-positive chronic infection, HBeAg-positive chronic hepatitis, HBeAg-negative chronic infection, HBeAg-negative chronic hepatitis, and HBsAg-negative phase, respectively.

[b]This patient had history of chronic HBV infection and did not receive antiviral therapy, but showed negative HBsAg and positive anti-HBs at the enrollment.

were also comparably high. In addition, these patients had normal ALT levels (data not shown). Thus, all patients in group one were in the HBeAg-positive chronic infection phase, also known as the immune tolerant phase according to 2017 European association for the study of the liver (EASL) guidelines (EASL 2017).

Table 2 shows the virological characteristics in patients in group two at a single time point. Of the six index patients, three were over 60 years old and three others were over 40 years old. Most of the patients were in natural phases 2–4 based on the EASL guideline (EASL 2017). Noticeably, seven patients underwent spontaneous HBeAg seroconversion, two had seroclearance of HBeAg without development of anti-HBe, and five showed coexistence of HBeAg and anti-HBe. Additionally, one index patient (XIIa) spontaneously cleared HBsAg and developed anti-HBs with undetectable HBV DNA and normal ALT value, indicative of recovery from HBV infection, thus falling into the HBsAg-negative phase (EASL 2017).

High Conservation of *Pre-C/C* Gene in Young Mothers and Their Children During 6.0–7.2 Years

Comparison of *pre-C/C* gene (624 bp) sequences recovered during the second trimester and 5.5–6.7 years postpartum revealed that *pre-C/C* gene had no mutation in seven women, and one single nucleotide substitution in one woman (XIIIa) (Table 3). The sequences recovered from their children were identical to the sequences in their mothers during pregnancy. Moreover, among four reference strains (GQ205441, KR013798, KU519422, LC170476) that were isolated in China or Japan, the strain (GQ205441) in a neighboring city is evolutionarily closest to the sequences in group one (Fig. 1A). The variation rate of nucleotides between GQ205441 and the sequences in group one was only 0.48%–1.44%, and none leads to an amino acid variation (Table 3).

Divergent Mutations of *Pre-C/C* Gene in Older Spreader and the Adult Offspring

HBV DNA were undetectable in the serum samples from two index patients in group two, presumably due to the extremely low viral load. In 13 other patients in group two, *pre-C/C* sequences varied significantly among the patients in each family. Compared to those in the index patients or between other family members, the nucleotide sequences and amino acid residues in *pre-C/C* region showed 2.24%–5.61% and 4.8%–13.46% differences respectively (Table 3), significantly higher than those observed in group one ($P < 0.001$). When compared to GQ205441, the nucleotide sequences and amino acid residues in patient IXc, who was in phase 1, had 0.96% and 0.5% difference respectively, similar to those observed in group one;

Table 3 Nucleotide and amino acid variations in pre-C/C gene among family members or compared to GQ205441.

Family	Patient	Relation	Sex	Age (years)	Compared to index[a]		Compared to GQ205441	
					nt (%)	aa (%)	nt (%)	aa (%)
Group one								
I	Ia	Index	F	30.7	–	–	3 (0.48)	0 (0.0)
				37.0	0 (0.0)	0 (0.0)	3 (0.48)	0 (0.0)
	Ib	Daughter	F	5.9	0 (0.0)	0 (0.0)	3 (0.48)	0 (0.0)
II	IIa	Index	F	21.4	–	–	5 (0.8)	0 (0.0)
				28.0	0 (0.0)	0 (0.0)	5 (0.8)	0 (0.0)
	IIb	Daughter	F	6.1	0 (0.0)	0 (0.0)	5 (0.8)	0 (0.0)
III	IIIa	Index	F	25.5	–	–	4 (0.64)	0 (0.0)
				32.3	0 (0.0)	0 (0.0)	4 (0.64)	0 (0.0)
	IIIb	Daughter	F	6.4	0 (0.0)	0 (0.0)	4 (0.64)	0 (0.0)
IV	IVa	Index	F	25.8	–	–	4 (0.64)	0 (0.0)
				33.0	0 (0.0)	0 (0.0)	4 (0.64)	0 (0.0)
	IVb	Son	M	6.7	0 (0.0)	0 (0.0)	4 (0.64)	0 (0.0)
V	Va	Index	F	25.6	–	–	5 (0.8)	0 (0.0)
				32.7	0 (0.0)	0 (0.0)	5 (0.8)	0 (0.0)
	Vb	Son	M	6.7	0 (0.0)	0 (0.0)	5 (0.8)	0 (0.0)
VI	VIa	Index	F	22.5	–	–	8 (1.28)	0 (0.0)
				28.5	0 (0.0)	0 (0.0)	8 (1.28)	0 (0.0)
	VIb	Daughter	F	5.6	0 (0.0)	0 (0.0)	8 (1.28)	0 (0.0)
VII	VIIa	Index	F	27.4	–	–	4 (0.64)	0 (0.0)
				34.3	0 (0.0)	0 (0.0)	4 (0.64)	0 (0.0)
	VIIb	Daughter	F	6.5	0 (0.0)	0 (0.0)	4 (0.64)	0 (0.0)
VIII	VIIIa	Index	F	24.2	–	–	9 (1.44)	0 (0.0)
				30.2	1 (0.001)	0 (0.0)	8 (1.28)	0 (0.0)
	VIIIb	Son	M	5.5	0 (0.0)	0 (0.0)	9 (1.44)	0 (0.0)
Group two								
IX	IXa	Index	F	78	–	–	23 (3.69)	15 (7.2)
	IXb	Daughter	F	53	35 (5.61)	28 (13.46)	24 (3.85)	17 (8.2)
	IXc	Son	M	45	22 (3.53)	14 (6.73)	6 (0.96)	1 (0.5)
X	Xa	Index	F	71	–	–	9 (1.44)	5 (2.4)
	Xb	Son	M	40	14 (2.24)	10 (4.80)	13 (2.08)	7 (3.4)
XI	XIa	Index	F	63	Undetectable			
	XIb	Son	M	41	–	–	6 (0.96)	4 (1.9)
	XIc	Son	M	38	19 (3.04)	11 (5.29)	23 (3.69)	15 (7.2)
XII	XIIa	Index	F	46	Undetectable			
	XIIb	Daughter	F	23	–	–	14 (2.24)	10 (4.8)
	XIIc	Son	M	21	20 (3.21)	15 (7.21)	13 (2.08)	8 (3.8)
XIII	XIIIa	Index	M	46	–	–	24 (3.85)	14 (6.7)
	XIIIb	Son	M	22	27 (4.33)	19 (9.13)	7 (1.12)	7 (3.4)
XIV	XIVa	Index	F	40	–	–	15 (2.4)	10 (4.8)
	XIVb	Husband	M	40	16 (2.56)	12 (5.76)	13 (2.08)	7 (3.4)

[a]Pre-C/C sequences were compared to the sequences recovered from the index patients during the second trimester in families I–VIII, and to the sequences recovered from the index patients or other infected family members in families IX–XIV.

whereas the nucleotide sequences and amino acid residues in the other 12 patients had 1.12%–3.68% and 1.9%–8.2% difference respectively, significantly higher than those observed in group one ($P < 0.001$). Phylogenetic analysis showed highly close evolutionary relationship between each mother (both during pregnancy and at 5.5–6.7 years

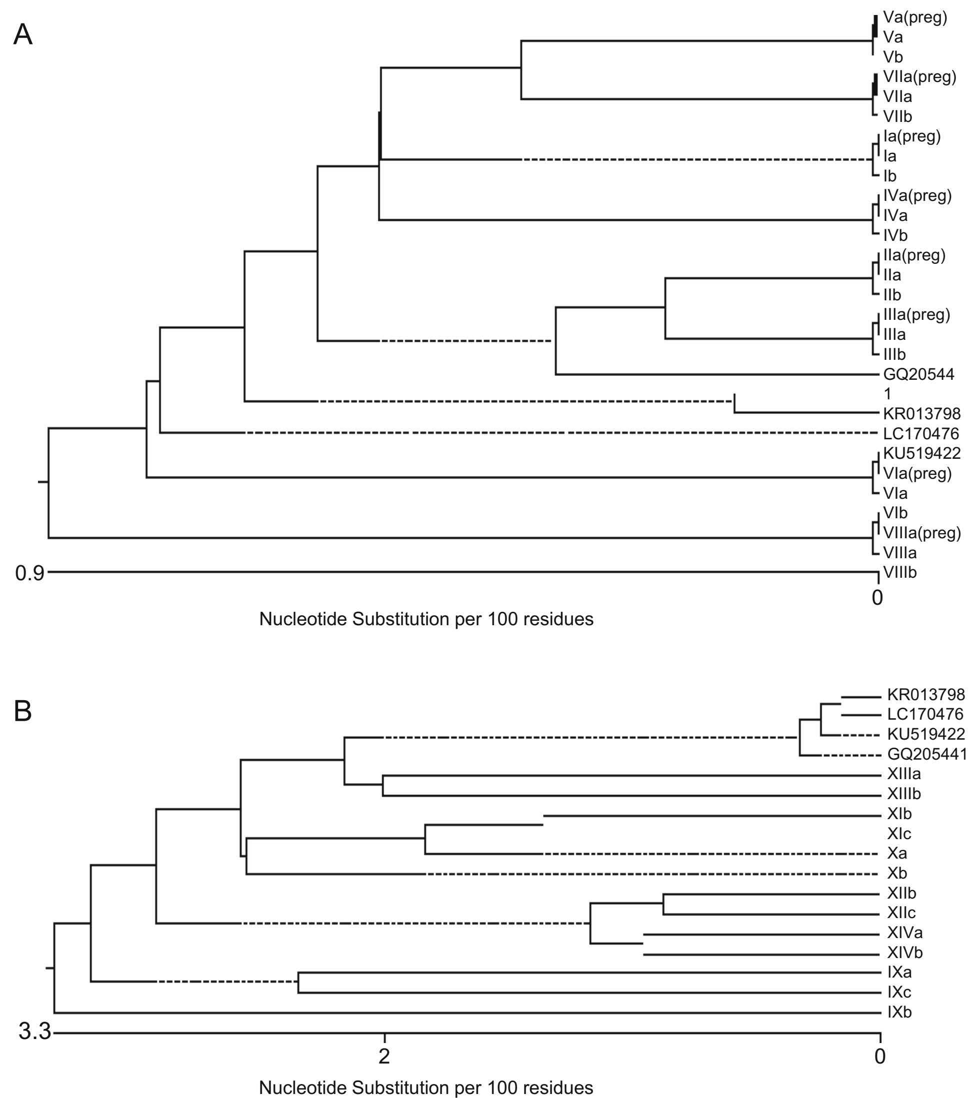

Fig. 1 Phylogenetic analysis of *pre-C/C* region of HBV. Phylogenetic tree was constructed based on the *pre-C/C* sequences (624 bp) from samples in families I–VIII in group one (**A**) and families IX–XIV in group two (**B**). The sequences GQ205441 (Heifei, China), KR013798 (Guangzhou, China), KU519422 (Tibet, China), LC170476 (Tokyo, Japan) (all genotype C) retrieved from GenBank were used as references.

postpartum) and her child in all families in group one (Fig. 1A), but relatively distant relationship among members in each family in group two (Fig. 1B).

Non-synonymous Substitutions Spread Out Immune Epitopes in Core Protein and Prevail in G1896

When compared with GQ205441, a total of 106 different point substitutions were detected in *pre-C/C* gene in this study. No deletions or insertions were found within *pre-C/C* gene. Of them, 75 nucleotide substitutions, including six double mutations and two triple mutations, lead to 64 missense mutations and one nonsense mutation (i.c. G1896A). Sixty-one amino acid variations occurred in core protein (Fig. 2), the majority (72%) of which spread out within previously reported epitopes for T cells, cytotoxic T lymphocytes (CTLs), or B cells (Salfeld *et al.* 1989; Bertoletti *et al.* 1991, 1993; Ferrari *et al.* 1991; Sallberg *et al.* 1993; Carman *et al.* 1997). The other four non-synonymous mutations occurred in *pre-C* gene, among which the most prevalent one is the nonsense mutation G1896A, found in five HBeAg-negative patients (XIc, IXb, Xa, Xb, XIVb) and three (IXa, XIb, XIVa) who was undergoing seroconversion.

Discussion

In the present study, we investigated the evolution of *pre-C/C* gene sequences in 31 patients from 14 families with known perinatal/intrafamilial HBV infection. We found that in the immune tolerant phase, *pre-C/C* sequences remained almost unchanged during 6.0–7.2 years in both the spreaders and the infected children, and had high homology to the sequences of HBV strain GQ205441

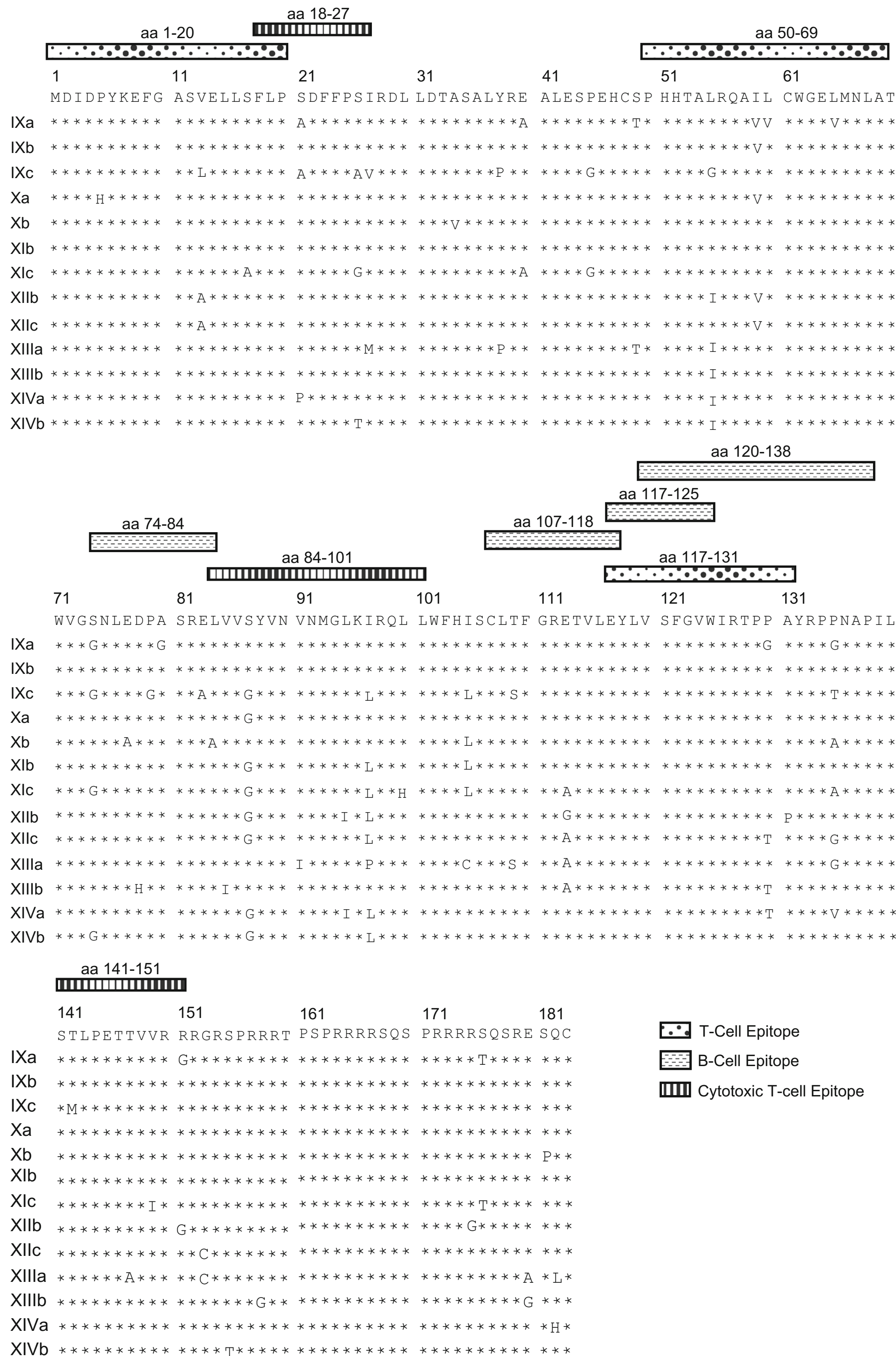

Fig. 2 All sixty-one missense mutations found in core protein in this study. Missense mutations were found in amino acids (aa) of core protein in each sample from families IX–XIV, the majority of which spread out known immune epitopes for T cells, CTLs, or B cells.

isolated from a nearby city (Zhang *et al.* 2011). By contrast, the intrafamilial transmitted HBV had gained diversified mutations in *pre-C/C* gene among the family members mostly in immune-clearance and low replicative phases. These results indicate that HBV *pre-C/C* gene is highly conservative during the immune tolerant phase, and mutations emerge around the time of the immune-clearance phase that is often coincided with decrease or loss of HBeAg. Seemingly, it may be the host immune pressure that decidedly drives the genesis of HBV mutations.

It has long been considered that the viral reverse transcriptase of HBV has an error-prone nature, and HBV thus evolves with a high mutation rate similar to that of RNA viruses (Holmes 2008). The estimated half-life of circulating HBV varies, from 2.5 to 46 min (Dandri *et al.* 2008) to 4–24 h (Murray *et al.* 2005, 2006). In the present study, HBV DNA levels in the mothers from group one during the second trimester were comparable to those at 5.5–6.7 years postpartum (Table 1). To maintain such constant circulating viral loads, the circulating HBV should be replenished at least by 50% per day based on an estimated half-life of 24 h. By simple calculation, the HBV in the paired mother–child should have noticeable mutations during the observation period of 6.0–7.2 years (Girones and Miller 1989). However, of eight women, seven did not have any substitution, and one had only one nucleotide substitution yet nonsense mutation at 6.0–7.2 years follow-up (Table 3). Moreover, the *pre-C/C* gene in the eight children had identical sequences to their mothers (Table 3, Fig. 1A). These results indicate that in the immune tolerant phase, even the error-prone nature of HBV replication was hardly able to introduce detectable mutation into *pre-C/C* gene.

Noticeably, compared with the sequence of HBV GQ205441, which was isolated in a nearby city in China, the *pre-C/C* sequences from individuals in immune tolerant phase (group one) had high (98.56%–99.52%) homology at nucleotide level and complete homology (100%) at amino acid level (Table 3). The patient (IXc) in immune tolerant phase in group two also had high homology at nucleotide and amino acid levels with GQ205441 (Table 3). The high homology was unlikely caused by cross-contamination, as GQ205441 was isolated in a different institute, and the sequences among all the index patients were indeed diversified (Fig. 1). These findings suggest that a predominant HBV isolate is circulating in this region, and its *pre-C/C* gene has maintained highly conservative in immune tolerant phase even among different individuals. These results also imply that with little host immune pressure in play, proof-reading-deficient viral reverse transcriptase rarely causes detectable replication error.

On the other hand, the *pre-C/C* gene in the subjects who were in non-immune tolerant phases (group two) showed diversified sequences (Table 3, Fig. 1B), despite that the family members most likely acquired the same HBV sources from the index patients. Of the 15 patients, one was in the immune tolerant phase with high viral load over 1.0×10^{8}, and 14 others were in other phases among which 10 had significantly reduced HBV DNA levels of $\sim 1.0 \times 10^{-4}$ IU/mL (Table 2). Thus, it is probable that most patients in group two had developed specific immune pressure against HBV, at least during a certain period. Therefore, the *pre-C/C* gene mutations were likely resulted from the long-term immune pressure as reported previously (Wang *et al.* 2010).

The underlying mechanism that drives HBV evolution during its chronic infection has intrigued microbiologists for long (Croagh and Lubel 2014; Warner *et al.* 2014; Boeijen *et al.* 2017; Faure-Dupuy *et al.* 2017; Lazarevic *et al.* 2019). A quasi-species theory has been proposed to explain the interplay between the host immune pressure and the HBV genomic diversity (Warner *et al.* 2014). When there is only low immune pressure (such as in the immune tolerant phase), high viral load is maintained, whereas a low selective pressure exists that leads to only few emerging adaptive mutants. Positive selection pressure occurs (such as in the immune-clearance phase) when host immunity that curtails HBV replication forces the selection of virions that contain escape mutations in the immune epitopes they recognize. Mutations under positive selection can be identified in viral subcloning by finding a high rate of non-synonymous mutations in the genes that encode the immune epitopes. An increased frequency of positively selected mutations has been shown in the *pre-C/C* gene of the HBV from HBeAg-negative patients (Abbott *et al.* 2010). In this study, samples from 9 out of 13 patients in group two required subcloning to determine the correct sequences due to mixed trace signals in the chromatogram after direct sequencing (data not shown). For each sample, a consensus sequence was determined and used. Whereas samples from group one showed rival diversity to much less extent that only one sample needed subcloning and showed variants with only synonymous mutations in *pre-C/C* gene. In consistent, a previous longitudinal study spanning 4–14 years that followed-up 18 patients who were treatment naïve showed that the viral diversity may stay very low and stable for many years during immune tolerant phase, followed by an increase in the viral diversity within 0–3 years around the time of HBeAg seroconversion (Nie *et al.* 2012). To reveal more detailed viral mutational spectrums in patients, next generation sequencing (NGS) may be required in further in-depth analysis. Indeed, NGS technologies are able to assess the mutational frequency per site and have revolutionized the way to study diversity of viral population.

HBV is a non-cytopathic virus, and the infection itself does not damage hepatocytes. Liver damage arises from cytolytic effects of the immune system, mainly the CTLs which attempt to clear HBV by killing the infected cells (Maini *et al.* 2000). The magnitude of such immune response has been noted to determine the course of the infection and clinical outcomes (Boeijen *et al.* 2017). Vigorous immune attack against infection is evidenced by elevated ALT (which are released from injured or killed hepatocytes) with suppressed viral replication. While normal ALT and active viral replication are signs for host immunity tolerant to HBV infection. This is the rationale behind the phase classification of HBV infection. Although we did not directly examine the CTL immunity in these patients, subcloning results revealed variants in group two harboring a dozen of different amino acid changes spread out immune epitopes in the core protein, including those for $CD8^+$ T cells, $CD4^+$ T cells, and B cells, co-exist in the serum. Clustering mutations occurred in known immunological target regions may reflect a role of the immune response for selection. On the other hand, no conclusive evidence has so far demonstrated that the accumulation of *pre-C/C* mutations could influence the outcomes of liver diseases.

There are several limitations in this study. First, the blood samples in the eight children in group one were available only at 5.5–6.7 years age, leaving us impossible to observe the evolution of viral sequences in these children. However, the full sequence homology between the children and their mothers indicated that the sequences in children did not undergo mutation. Second, the samples in all individuals in group two were cross-sectional. Thus, we could not longitudinally observe the evolution of the same original HBV during the different phases of infection. Third, we cannot exclude potential mutations in regions other than *pre-C/C*. In fact, *pre-C/C* encodes HBeAg and HBcAg, two major targets for CTLs- and B cells-mediated antiviral immunity. HBV e and c antigens are prone to a wide variation of amino acid sequences during chronic HBV infection (Wang *et al.* 2016; Colombatto *et al.* 2018; Luo *et al.* 2019). The *pre-C/C* is arguably the most common site for mutations in HBV (Wang *et al.* 2016; Luo *et al.* 2019). Mutations occurred to *pre-C/C* should reflect how conservative or radical the virus had evolved so far. Fourth, the number of study subjects was relatively small. Nevertheless, these limitations did not taint the finding that *pre-C/C* is highly conservative for 6.0–7.2 years in the immune tolerant phase regardless of individuals.

In summary, HBV *pre-C/C* gene in perinatally HBV-infected patients is highly conservative during the immune tolerant phase irrespective of different individuals, and undergoes diversified mutations around the time of ongoing immune-clearance phase. These results indicate that the emergence of viral variants is mainly driven by the host immune pressure. Mutations occurred in *pre-C/C* gene may mark the switch of immune tolerant to immune reactive phase in the natural course of HBV infection.

Acknowledgements We thank Ms. Zhenhua Feng (Nanjing Drum Tower Hospital, Nanjing 210008, China) for performing sequencing of HBV *pre-C/C* gene. This study was supported by the National Natural Science Foundation of China (81672002), the Science and Technology Department of Jiangsu Province (BK20161105), and the Jiangsu Provincial Department of Health (H201537), China.

Author Contributions YHZ designed the study and critically revised the manuscript; YL and LZ performed the experiments, analyzed data and wrote the manuscript; YD and YH followed the patients, collected and analyzed the clinical data; BX participated in the design and performed the statistical analysis. All authors read and approved the final manuscript.

Compliance with Ethical Standards

Conflict of interest The authors declare no conflict of interests.

Animal and Human Rights Statement This study was approved by the institutional review boards of Nanjing Drum Tower Hospital, essentially following the ethical guidelines of the Declaration of Helsinki. Before sample collection, written informed consent was obtained from all the patients or their guardians for which identifying information is included in this article.

References

Abbott WG, Tsai P, Leung E, Trevarton A, Ofanoa M, Hornell J, Gane EJ, Munn SR, Rodrigo AG (2010) Associations between HLA class I alleles and escape mutations in the hepatitis B virus core gene in New Zealand-resident Tongans. J Virol 84:621–629

Bertoletti A, Ferrari C, Fiaccadori F, Penna A, Margolskee R, Schlicht HJ, Fowler P, Guilhot S, Chisari FV (1991) HLA class I-restricted human cytotoxic T cells recognize endogenously synthesized hepatitis B virus nucleocapsid antigen. Proc Natl Acad Sci USA 88:10445–10449

Bertoletti A, Chisari FV, Penna A, Guilhot S, Galati L, Missale G, Fowler P, Schlicht HJ, Vitiello A, Chesnut RC (1993) Definition of a minimal optimal cytotoxic T-cell epitope within the hepatitis B virus nucleocapsid protein. J Virol 67:2376–2380

Boeijen LL, Hoogeveen RC, Boonstra A, Lauer GM (2017) Hepatitis B virus infection and the immune response: the big questions. Best Pract Res Clin Gastroenterol 31:265–272

Carman WF, Thursz M, Hadziyannis S, McIntyre G, Colman K, Gioustoz A, Fattovich G, Alberti A, Thomas HC (1995) Hepatitis B e antigen negative chronic active hepatitis: hepatitis B virus core mutations occur predominantly in known antigenic determinants. J Viral Hepat 2:77–84

Carman WF, Boner W, Fattovich G, Colman K, Dornan ES, Thursz M, Hadziyannis S (1997) Hepatitis B virus core protein mutations are concentrated in B cell epitopes in progressive disease and in T helper cell epitopes during clinical remission. J Infect Dis 175:1093–1100

Clewley JP, Arnold C (1997) MEGALIGN. The multiple alignment module of LASERGENE. Methods Mol Biol 70:119–129

Colombatto P, Barbera C, Bortolotti F, Maina AM, Moriconi F, Cavallone D, Calvo P, Oliveri F, Bonino F, Brunetto MR (2018) HBV pre-core mutant in genotype-D infected children is selected during HBeAg/anti-HBe seroconversion and leads to HBeAg negative chronic hepatitis B in adulthood. J Med Virol 90:1232–1239

Croagh CM, Lubel JS (2014) Natural history of chronic hepatitis B: phases in a complex relationship. World J Gastroenterol 20:10395–10404

Dandri M, Murray JM, Lutgehetmann M, Volz T, Lohse AW, Petersen J (2008) Virion half-life in chronic hepatitis B infection is strongly correlated with levels of viremia. Hepatology 48:1079–1086

European Association for the Study of the Liver (EASL) (2017) EASL 2017 Clinical Practice Guidelines on the management of hepatitis B virus infection. J Hepatol 67:370–398

Faure-Dupuy S, Lucifora J, Durantel D (2017) Interplay between the hepatitis B virus and innate immunity: from an understanding to the development of therapeutic concepts. Viruses 9:95

Ferrari C, Bertoletti A, Penna A, Cavalli A, Valli A, Missale G, Pilli M, Fowler P, Giuberti T, Chisari FV et al (1991) Identification of immunodominant T cell epitopes of the hepatitis B virus nucleocapsid antigen. J Clin Invest 88:214–222

Girones R, Miller RH (1989) Mutation rate of the hepadnavirus genome. Virology 170:595–597

Holmes EC (2008) Evolutionary history and phylogeography of human viruses. Annu Rev Microbiol 62:307–328

Hu Y, Zhang S, Luo C, Liu Q, Zhou YH (2012) Gaps in the prevention of perinatal transmission of hepatitis B virus between recommendations and routine practices in a highly endemic region: a provincial population-based study in China. BMC Infect Dis 12:221

Kimura M (1980) A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J Mol Evol 16:111–120

Lazarevic I, Banko A, Miljanovic D, Cupic M (2019) Immune-escape hepatitis B virus mutations associated with viral reactivation upon immunosuppression. Viruses 11:778

Liang X, Cheng J, Sun Y, Chen X, Li T, Wang H, Jiang J, Chen X, Long H, Tang H, Yu Y, Sheng J, Chen S, Niu J, Ren H, Shi J, Dou X, Wan M, Jiang J, Xie Q, Shi G, Ning Q, Chen C, Tan D, Ma H, Sun J, Jia J, Zhuang H, Hou J (2015) Randomized, three-arm study to optimize lamivudine efficacy in hepatitis B e antigen-positive chronic hepatitis B patients. J Gastroenterol Hepatol 30:748–755

Liu J, Bi Y, Xu C, Liu L, Xu B, Chen T, Chen J, Pan M, Hu Y, Zhou YH (2015) Kinetic changes of viremia and viral antigens of hepatitis B virus during and after pregnancy. Medicine 94:e2001

Luo Y, Pan M, Ning M, Chenyu X, Liu L, Chen L, Chen T, Biyun X, Yali H, Zhou YH (2019) High mutation prevalence of precore and basal core promoter in pregnant women who underwent spontaneous HBeAg seroconversion within one year postpartum. Dig Liver Dis 52:199–204

Maini MK, Boni C, Lee CK, Larrubia JR, Reignat S, Ogg GS, King AS, Herberg J, Gilson R, Alisa A, Williams R, Vergani D, Naoumov NV, Ferrari C, Bertoletti A (2000) The role of virus-specific CD8(+) cells in liver damage and viral control during persistent hepatitis B virus infection. J Exp Med 191:1269–1280

Murray JM, Wieland SF, Purcell RH, Chisari FV (2005) Dynamics of hepatitis B virus clearance in chimpanzees. Proc Natl Acad Sci USA 102:17780–17785

Murray JM, Purcell RH, Wieland SF (2006) The half-life of hepatitis B virions. Hepatology 44:1117–1121

Nie H, Evans AA, London WT, Block TM, Ren XD (2012) Quantitative dynamics of hepatitis B basal core promoter and precore mutants before and after HBeAg seroconversion. J Hepatol 56:795–802

Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406–425

Salfeld J, Pfaff E, Noah M, Schaller H (1989) Antigenic determinants and functional domains in core antigen and e antigen from hepatitis B virus. J Virol 63:798–808

Sallberg M, Pushko P, Berzinsh I, Bichko V, Sillekens P, Noah M, Pumpens P, Grens E, Wahren B, Magnius LO (1993) Immunochemical structure of the carboxy-terminal part of hepatitis B e antigen: identification of internal and surface-exposed sequences. J Gen Virol 74(Pt 7):1335–1340

Tamura K, Nei M, Kumar S (2004) Prospects for inferring very large phylogenies by using the neighbor-joining method. Proc Natl Acad Sci USA 101:11030–11035

Valaydon ZS, Locarnini SA (2017) The virological aspects of hepatitis B. Best Pract Res Clin Gastroenterol 31:257–264

Vanlandschoot P, Cao T, Leroux-Roels G (2003) The nucleocapsid of the hepatitis B virus: a remarkable immunogenic structure. Antiviral Res 60:67–74

Wang HY, Chien MH, Huang HP, Chang HC, Wu CC, Chen PJ, Chang MH, Chen DS (2010) Distinct hepatitis B virus dynamics in the immunotolerant and early immunoclearance phases. J Virol 84:3454–3463

Wang XL, Ren JP, Wang XQ, Wang XH, Yang SF, Xiong Y (2016) Mutations in pre-core and basic core promoter regions of hepatitis B virus in chronic hepatitis B patients. World J Gastroenterol 22:3268–3274

Warner BG, Abbott WG, Rodrigo AG (2014) Frequency-dependent selection drives HBeAg seroconversion in chronic hepatitis B virus infection. Evol Med Public Health 2014:1–9

Xu B, Zhu D, Bi Y, Wang Y, Hu Y, Zhou YH (2017) Minimal association of alleles of human leukocyte antigen class II gene and long-term antibody response to hepatitis B vaccine vaccinated during infancy. Vaccine 35:2457–2462

Zhang ZH, Li L, Zhao XP, Glebe D, Bremer CM, Zhang ZM, Tian YJ, Wang BJ, Yang Y, Gerlich W, Roggendorf M, Li X, Lu M, Yang DL (2011) Elimination of hepatitis B virus surface antigen and appearance of neutralizing antibodies in chronically infected patients without viral clearance. J Viral Hepat 18:424–433